Report for Research Project

"A New Framework for Large Scale Data Envelopment Analysis (DEA)."

Proposal Number:

ONR 98342-0080

Principal Investigator:

José H. Dulá

Associate Professor

Department of Management and Marketing

e-mail: jdula@olemiss.edu

1. Summary

The research agenda presented for this project has been completed. This includes the following tasks:

- A new procedure for large scale DEA computations based on the identification of the essential elements (the *frame*) of the data set.

- Testing of the codes on randomly generated data. The final results indicate that the procedure substantially reduces computational times and imparts a new level of flexibility on DEA analyses. Figure 1 (p.11) of the final paper shows how, when the density of efficient DMUs is low, it is possible to perform studies using the four standard DEA models for a tiny fraction of what it costs to perform these studies separately, as is currently the standard approach.

- Testing of the codes on Navy data. The procedure was tested on data extracted from one of the Navy's master EMF file, 'EMF.9711.' A total of four implementations were performed the largest consisting of a study involving 10,529 DMUs (enlisted men) and 11 inputs & outputs.

The following scholarly activities were carried out or are planned connected with this research project:

- Presentations:

1. Dulá, J.H. and R.M. Thrall, "Accelerating DEA over multiple models & forms with frames," INFORMS Cincinnati, May 2-5, 1999, Cincinnati, OH.

2. Dulá, J.H. and R.M. Thrall, "Accelerating DEA over multiple models and forms with frames." 5th International Conference of the Decision Sciences Institute, Athens, Greece July 4-7, 1999.

19990607 012

- Papers:

  1. Dulá, J.H. and R.M. Thrall, "A computational framework for accelerating DEA." Submitted to *J. Productivity Analysis* (attached).

<u>Remarks.</u> The paper "A computational framework for accelerating DEA," submitted to *JPA* is a finalist in the 'Best Paper Competition, Theory' for the DSI conference in Athens, Greece, this coming July. The nomination came from Prof. W.W. Cooper from the University of Texas at Austin.

The manuscript "A computational framework for accelerating DEA" attached to this report contains the results of the research and its application to randomly generated data. Below we will present a report of the results using Navy data.

2. Application on Navy data.

In this research project, a methodology and a code were developed to apply DEA to large data sets. A DEA implementation is considered 'large scale' when the number of units (DMUs) and/or the number of dimensions (inputs plus outputs) is relatively large. This definition, of course, is continually changing as computing resources continue to evolve rapidly. Currently, most DEA implementations involve fewer than 1000 DMUs in under ten dimensions. These problems can be solved in reasonable time on an ordinary personal computer using macros and solvers in commercial spreadsheets. Implementations with between 1000 and 5000 DMUs would be modestly large and would require specialized software. Implementations with more than 10,000 DMUs and more than, say, ten dimensions, are truly large scale and are relatively rare. Such large scale applications are impractical to solve in a normal desktop computer.

The methodology and program developed for this project was applied to a problem using Navy data. A DEA model was built based on the availability of data in the Navy's EMF files. The model was constructed to assess efficiency based on how Navy personnel transform practical skills, intellectual ability, educational background, and accumulated experience into performance levels that the Navy evaluates and tracks individually. Skills, intellectual ability, educational background, and experience of an individual are considered inputs in the sense that they are "assets" or "endowments" which the individual applies toward achieving potential performance outputs. The objective is to identify those who attain their potential or exceed expectations. We naturally expect higher outputs from individuals who demonstrate intellectual abilities and possess higher levels of skill, education, training, and experience and would consider individuals with similarly high levels in their assets who attain less as inefficient. Conversely, individuals who are not particularly well endowed and who perform at unexpectedly high levels would be classified as efficient.

Such individuals may be of particular interest to the Navy as worthy of distinction and reward. The DEA model is designed to detect efficient and inefficient individuals as defined by their ability to transform their endowment into performance.

The DEA model used to measure performance by the standards described above used measures of experience, education and intellectual ability as inputs and standard evaluation scores in diverse categories as outputs. All input and output values were from ordinal scales. The initial extraction of data was from the file EMF.9711 and was limited to records in the E8 and E9 paygrade categories. The extraction yielded an initial total of 13,354 records. ¿From these, three categories were identified based on the type of tests appearing under the heading 'TEST-ID3.' They are

- Basic Test Battery (BTB).

- Armed Services Vocational Aptitutude Battery (ASVAB) Series 5-7.

- Armed Services Vocational Aptitutude Battery (ASVAB) Series 8-22.

For the BTB category the following were used as inputs for the DEA model:

1. 'LOS' — Length of Service (as calculated by T. Blackstone)

2. 'ED YEARS' — Years of education:
   Entry "ED-YRS" (Ch.3-1, col. 0331 in EMF.9711)

3. 'AFQT' — Armed Forces Qualification Test Score:
   Entry "AFQT-SCORE" (Ch. 3-19, col. 2338 in EMF.9711)

4. 'GEN CLASS' — General Classification Test Score:
   Entry "GCT" (Ch. 3-23, col. 2302 in EMF.9711).

5. 'ARITHMETIC' — Arithmetic Test Score:
   Entry "ARI" (Ch. 3-24, col. 2304 in EMF.9711)

6. 'MECHANICAL' — Mechanical Test Score:
   Entry "MEC"' (Ch. 3-25, col.2306 in EMF.9711).

7. 'CLERICAL' — Clerical Aptitude Test Score:
   Entry "CLER" (Ch. 3-26, col.2308 in EMF.9711).

For the ASVAB Series 5-7 category the following were used as inputs for the DEA model:

1. 'LOS' — Length of Service (as calculated by T. Blackstone)

2. 'ED YEARS' — Years of education:
   Entry "ED-YRS" (Ch.3-1, col. 0331 in EMF.9711)

3. 'AFQT' — Armed Forces Qualification Test Score:
   Entry "AFQT-SCORE" (Ch. 3-19, col. 2338 in EMF.9711)

4. 'GEN-INFO' — General Information Test:

Entry "GEN-INFO" (Ch. 3-28, col. 2302 in EMF.9711).

5. 'NUM-OPS'   Numerical Operations Test Score:

Entry "NUM-OPS" (Ch. 3-29, col. 2304 in EMF.9711)

6. 'ATTN-DETAIL'   Attention to Detail Test Score:

Entry "ATTN-DETAIL" (Ch. 3-30, col.2306 in EMF.9711).

7. 'WORD KNOW'   Word Knowledge Test Score:

Entry "WORD KNOW" (Ch. 3-31, col.2308 in EMF.9711).

8. 'ARITH'   Arithmetic Reasoning Test Score:

Entry "ARI-REAS" (Ch. 3-32, col.2310 in EMF.9711).

9. 'SPACE'   Space Perception Test Score:

Entry "SPACE-PERCEP" (Ch. 3-33, col.2312 in EMF.9711).

10. 'MATH'   Mathematical Knowledge Test Score:

Entry "MATH KNOW" (Ch. 3-34, col.2314 in EMF.9711).

11. 'ELECTRONIC'   Electronics Information Test Score:

Entry "ELEC-INFO" (Ch. 3-35, col.2316 in EMF.9711).

12. 'MECHANIC'   Mechanical Comprehension Test Score:

Entry "MECH-COMP" (Ch. 3-36, col.2318 in EMF.9711).

13. 'GEN-SCIENCE'   General Science Test Score:

Entry "GEN-SCI" (Ch. 3-37, col.2320 in EMF.9711).

14. 'SHOP'   Shop Information Test Score:

Entry "SHOP-INFO" (Ch. 3-38, col.2322 in EMF.9711).

15. 'AUTO'   Automotive Information Test Score:

Entry "AUTO-INFO" (Ch. 3-39, col.2324 in EMF.9711).

For the ASVAB Series 8-22 category the following were used as inputs for the DEA model:

1. 'LOS'   Length of Service (as calculated by T. Blackstone)

2. 'ED YEARS'   Years of Education:

Entry "ED-YRS" (Ch.3-1, col. 0331 in EMF.9711)

3. 'AFQT'   Armed Forces Qualification Test Score:

Entry "AFQT-SCORE" (Ch. 3-19, col. 2338 in EMF.9711)

4. 'GEN SCIENCE'   General Science Test:

Entry "GSC" (Ch. 3-40, col. 2302 in EMF.9711).

5. 'ARITHMETIC'   Arithmetic Reasoning Test:

Entry "ARR" (Ch. 3-41, col. 2304 in EMF.9711)

6. 'WORD'   Word Knowledge Test Score:

Entry "WOR" (Ch. 3-42, col.2306 in EMF.9711).

7. 'PARAGRAPH' Paragraph Comprehension Test Score:

Entry "PAR" (Ch. 3-43, col.2308 in EMF.9711).

8. 'NUM OPS' Numerical Operations Test Score:

Entry "NUM" (Ch. 3-44, col.2310 in EMF.9711).

9. 'CODING' Coding Speed Test Score:

Entry "COD" (Ch. 3-45, col.2312 in EMF.9711).

10. 'AUTO&SHOP' Auto and Shop Information Test Score:

Entry "ASI" (Ch. 3-46, col.2314 in EMF.9711).

11. 'MATH' Mathematics Knowledge Test Score:

Entry "MAT" (Ch. 3-47, col.2316 in EMF.9711).

12. 'MECHANICAL' Mechanical Comprehension Test Score:

Entry "MEC" (Ch. 3-48, col.2318 in EMF.9711).

13. 'ELECTRONIC' Electronic Information Test Score:

Entry "ELI" (Ch. 3-49, col.2320 in EMF.9711).

14. 'VERBAL' Verbal Test Score:

Entry "VER" (Ch. 3-50, col.2322 in EMF.9711).

A common set of outputs was found which was largely complete and could be used for all three categories. They were:

1. 'PROF KNOW' Evaluation Performance Traits: Professional Knowledge:

Entry "EVAL-PROF-KNOW" (Ch. 24-12, col.2979 in EMF.9711).

2. 'TEAMWORK' Evaluation Performance Traits: Teamwork:

Entry "EVAL-TEAM-WORK" (Ch. 24-13, col.2980 in EMF.9711).

3. 'LEADER' Evaluation Performance Traits: Leadership:

Entry "EVAL-LEADERSHIP" (Ch. 24-14, col.2981 in EMF.9711).

4. 'EQUAL OPP' Evaluation Performance Traits: Equal Opportunity:

Entry "EVAL-EQUAL-OPP" (Ch. 24-15, col.2982 in EMF.9711).

5. 'JOB ACC' Evaluation Performance Traits: Personal Job Accomplishment/Initiative:

Entry "EVAL-PERS-JOB-ACC" (Ch. 24-18, col.2985 in EMF.9711).

6. 'MISSION' Evaluation Performance Traits: Mission Accomplishment and Initiative:

Entry "EVAL-MISS-ACC" (Ch. 24-19, col.2986 in EMF.9711).

After processing the data and discarding invalid records (including records with input scores of zeroes) the following three data sets were created

Category name:	'BTB'	'ASVAB 5-7'	'ASVAB 8-22'
No. of Records:	3,821	4,870	1,837
Inputs:	7	15	14
Outputs:	6	6	6
Dimension (Total):	13	21	20

The three data sets were combined to test the procedure on a large data set. In order to make the union of the three data sets meaningful, a list of common inputs had to be found. Clearly, 'Length of Service,' 'Years of Education,' and 'Armed Forces Qualification Test Score' could be used in the combination. Other inputs were selected by finding parameters which could be thought as comparable across all three data sets. The final list of inputs for the data set containing the entire collection of records was:

1. 'LOS'
Length of Service (as calculated by T. Blackstone)

2. 'ED YEARS'
Years of Education:
Entry 'ED-YRS' (Ch.3-1, col. 0331 in EMF.9711)

3. 'AFQT'
Armed Forces Qualification Test Score:
Entry 'AFQT-SCORE' (Ch. 3-19, col. 2338 in EMF.9711)

4. 'ARITH'
Normalized score from "Arithmetic" type tests in all three categories:
Entry 'ARI' in BTB.
Entry 'ARI-REAS' in ASVAB Series 5-7.
Entry 'ARR' in ASVAB Series 8-22.

5. 'MECHANIC'
Normalized score from "Mechanic" type tests in all three category:
Entry 'MECH' in BTB.
Entry 'MECH-COMP' in ASVAB Series 5-7.
Entry 'MEC' in ASVAB Series 8-22.

The characteristics of the comprehensive model are:

Category:	'COMPREHENSIVE'
No. of Records:	10,529
Inputs:	5
Outputs:	6
Dimension (Total):	11

The three data sets, 'BTB,' 'ASVAB 5-7,' and 'ASVAB 8-22' are each under 5,000 DMUs. Since they all have more than ten dimensions, however, they can be considered large. Of special interest is 'ASVAB 5-7' because it has 21 dimensions. The number of dimensions has a much greater incremental impact on the computational burden than the number of DMUs. Therefore, 'ASVAB 5-7' being a model with 4,870 DMUs and 21 dimensions, offers a valuable opportunity to test the new code. Finally, it is well known in DEA computations using traditional approaches that, as the number of DMUs increases, the computational requirements become explosive. This tends to occur at some point between 5,000 and 10,000 DMUs. Our data set 'COMPREHENSIVE' is well beyond this point and makes and excellent test problem for any computational procedure.

The ideas of this project were implemented in the Fortran program 'ALLFRAMES' which was used to process the four data sets. The code first evaluates the *frame* of the data set for the "variable returns" (VR) DEA model. The frame of a DEA data set for a given model is the subset of DMUs which are "extreme-efficient." Extreme-efficiency is the main subset of the efficiency set. Only in rare cases this is not the full set. An important result is that extreme-efficient DMUs cannot be "weakly" efficient. This excludes the possibility of a difficult complication usually present in traditional DEA analysis. The procedure proceeds to find the other three frames corresponding to the increasing (IR), decreasing (DR) and constant (CR) returns models. The results of this analysis are given in Table 1.

Table 1.
Frame Cardinalities

Data Set	Number of DMUs	VR	IR	DR	CR
'BTB'	3,821	161	148	150	137
'ASVAB 5-7'	4,870	500	494	372	366
'ASVAB 8-22'	1,837	192	178	185	171
'COMPREHENSIVE'	10,529	101	96	82	77

This analysis identifies efficient individuals; that is, those whose input values tend to be low and performance scores high. To obtain a general profile of those who attained efficiency status let us look at their averages and compare them to the overall averages of the entire data sets. The following two tables are used for this comparison. Two tables are used to split the input and output components of the models. The report is for the variable returns (VR) model. First just inputs:

Table 2.
Input Averages: All DMUs vs Efficient

	INPUT*														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
'BTB' All DMUs	24.6	12.6	60.5	54.7	53.6	50.2	54.0	-	-	-	-	-	-	-	-
'BTB' Efficient	24.0	11.5	39.1	42.1	43.6	42.1	46.6	-	-	-	-	-	-	-	-
'ASVAB 5-7' All DMUs	19.6	12.4	66.0	53.9	54.1	52.7	55.9	55.9	54.1	56.8	55.9	53.8	55.8	54.1	53.5
'ASVAB 5-7' Efficient	18.8	12.0	48.1	44.2	49.8	50.0	49.7	49.7	48.1	50.6	48.7	46.6	48.7	47.3	45.5
'ASVAB 8-22' All DMUs	18.0	12.6	72.3	55.6	58.2	56.0	56.3	56.2	57.4	57.4	56.5	55.7	57.35	56.3	-
'ASVAB 8-22' Efficient	16.4	12.2	48.5	48.4	51.2	50.0	50.9	52.2	53.4	50.0	49.4	48.2	50.0	50.2	-
'COMPREHENSIVE' All DMUs	21.1	12.5	65.1	65.1	62.6	-	-	-	-	-	-	-	-	-	-
'COMPREHENSIVE' Efficient	19.6	11.2	34.6	32.2	38.4	-	-	-	-	-	-	-	-	-	-

*Refer to the preceding discussion for the input attribute titles corresponding to the indexes.

The important observation for Table 2 is that averages for the efficient subset are always better (lower for inputs) than for the entire data set. Table 3 presents the comparison of the common list of outputs. Here, note that the average for the efficient subset is higher than that of the entire data set for each input attribute.

Table 3.
Output Averages: All DMUs vs Efficient

	OUTPUT*					
	1	2	3	4	5	6
'BTB' All DMUs	4.3	4.4	4.0	4.0	4.4	4.1
'BTB' Efficient	4.5	4.6	4.2	4.4	4.6	4.4
'ASVAB 5-7' All DMUs	4.3	4.4	3.9	4.1	4.5	4.2
'ASVAB 5-7' Efficient	4.6	4.7	4.2	4.5	4.6	4.5
'ASVAB 8-22' All DMUs	4.3	4.4	3.8	4.2	4.5	4.2
'ASVAB 8-22' Efficient	4.5	4.6	4.1	4.4	4.6	4.4
'COMP' All DMUs	4.3	4.4	3.9	4.1	4.4	4.2
'COMP' Efficient	4.5	4.5	4.2	4.3	4.5	4.3

*Refer to the preceding discussion for the output attribute titles corresponding to the indexes.

All individuals identified in the analysis as 'efficient' are remarkable in some way. They combine a unique set of input and output attribute values which place them on the empirical efficient frontier corresponding to their data set. Some remarkable individuals are easily identified by the analysis. For example, consider the individual with index 13 in data set 'COMPREHENSIVE'†

Table 4.
Attributes of Individual "13" in
Data Sets 'BTB' and 'COMPREHENSIVE'

| | ATTRIBUTE | | | | | | |
	1	2	3	4	5	6	7
INPUT 'BTB'	25	10	14	35	32	29	29
INPUT 'COMPREHENSIVE'	25	10	14	15.1	14.6	-	-
OUTPUT	5	5	4	4	5	5	-

This DMU has emerged as efficient because it corresponds to an individual with relatively little education (10 yrs) and particularly low academic and vocational abilities but clearly able to attain relatively high evaluation scores. One possible assessment of this individual is that he/she may be compensating for deficiencies in education and basic skills with other personal qualities such as diligence and discipline.

Conversely, individuals who are classified as 'inefficient' would have relatively high input values and low evaluations. These would be individuals who, despite, experience, education and demonstrated abilities, are not able to present, through their actions and attitudes, performance levels that are rated highly.

¿From the analysis we are able to identify VR-efficient DMUs that demonstrate efficiency under other returns to scale assumptions. In the case of the same individual whose index was '13,' it is also the case that it is efficient under the constant returns to scale assumption. CR-efficiency is a more exclusive set than VR-efficiency. The constant returns model is more restrictive since an efficient DMU in this model dominates inefficient DMUs even if these are freely scaled uniformly up or down.

The data set was also processed using a specially coded DEA program applying the standard approach incorporating all known enhancements. The computational times of the new procedure of

---

† This individual can be identified via his/her SSN.

these computations appear in Table 5 below (times are in seconds on the University of Mississippi's central SGI computer).

Table ??.
Computation Times.

Data Set	Traditional (Enhanced) Secs.	Frames VR Secs.	IR+DR+CR Secs.[†]	Scoring Phase Secs.[‡]
'BTB'	289	14	1	6[‡]
'ASVAB 5-7'	977	96	20	44[‡]
'ASVAB 8-22'	85	10	2	5[‡]
'COMPREHENSIVE'	1,884	13	0.5	21.5

[†]Estimated (see Figure 1, p11, of "Accelerating DEA computations.")
[‡]Estimated.

The computational times for the new procedure are substantially less than for the traditional approach in all four data sets. Notice that this comparison includes finding four frames. We can expect time increases of four fold to evaluate the data set for all four models using the traditional approach. The procedure is affected by the dimension of the data set and density of the VR frame. When the dimension and density are low as in the data set "Comprehensive' (11 inputs and outputs and 1% VR frame density) the performance of the new procedure compared to traditional approaches is dramatically superior.

Conclusion. Evaluating and comparing efficiency and performance of many, functionally similar, units is an important part of the management of complex operations. For the Navy these are tasks that present themselves on a large scale. For example, the Navy must compare, measure, and evaluate individuals' activities within given ranks and functions. Additionally, the decision maker is interested in dynamic studies that track the progress of these individuals through time. This means repeatedly solving very large DEA models The decision maker needs effective quantitative tools that enhance his/her ability to analyze, measure, understand, and improve these aspects of the operation. Our application on Navy personnel validates the claim that the methodology is practical for very large data sets. It is reasonable to extrapolate, based on the results above, that DEA studies involving several tens of thousand DMUs are practical. It is conceivable that studies beyond 100 thousand DMUs are tractable. The Navy can now have the tools to perform evaluations and track performance for large subsets, and the entire set, of its corps of enlisted men.

José H. Dulá

Associate Professor of Production and Operations Management
School of Business
The University of Mississippi
University, MS 38677
Telephone: (601) 232-5473
e-mail: jdula@olemiss.edu

February 15, 1999

Prof. C.A. Knox Lovell,
Editor-in-Chief, *Journal of Productivity Analysis*
University of New South Wales
Sydney NSW 2052
AUSTRALIA

Dear Professor Lovell,

Please find enclosed four copies of the paper "A computational framework for accelerating DEA," which we are submitting for review in *Journal of Productivity Analysis*.

This paper formally establishes and successfully exploits a close interrelation among the four standard DEA models: "variable returns," "increasing returns," "decreasing returns," and "constant returns." The result is a framework for designing procedures using essential subsets of the data domain that, as we show, have the potential to accelerate substantially DEA computations when analyses are over these four models and multiple orientations. Our work makes DEA computations faster and adds flexibility. Its "orientation-free" aspect should interest researchers working on the impact of new and diverse envelopment objective functions in DEA analysis (e.g., Cooper, Park, and Pastor [1998], Thrall [1996b], Tone [1998], etc.).

If we may offer suggestions for editors, we would like to propose Prof. W.W. Cooper from the University of Texas at Austin. We feel this editor is ideally suited to handle the paper because of his understanding of linear programming and polyhedral set theory and because he is in a position to assess the contributions of the paper and select capable referees.

I thank you for your time and consideration and I look forward to your response.

Sincerely yours,

José H. Dulá

A computational framework for accelerating DEA.

J.H. Dulá[1]
The University of Mississippi
University, MS 38677

R.M. Thrall[2]
Rice University
Houston, TX 77251

February 1999

A computational framework for accelerating DEA.

ABSTRACT. We introduce a new computational framework for DEA that reduces computation times and increases flexibility in applications over multiple models and orientations. The process is based on the identification of *frames* —minimal subsets of the data needed to describe the problem– for each of the four standard production possibility sets. It exploits the fact that the frames of the models are closely interrelated. Access to a frame of a production possibility set permits a complete analysis in a second phase for the corresponding model either oriented or orientation-free. This second phase proceeds quickly especially if the frame is a small subset of the data points. The use of frames extends other tangible advantages such as a guarantee that a sufficient condition for weak efficiency will be verified in certain cases. Besides accelerating computations, the new framework imparts greater flexibility to the analysis by not committing the analyst to a model or orientation when performing the bulk of the calculations. Computational testing validates the results and reveals that, with a minimum additional time over what is required for a full DEA study for a given model and specified orientation, one can obtain the analysis for the four models and all orientations.

Key Words: DEA, DEA computations, linear programming, and convex analysis.

Introduction. DEA is a non-parametric frontier estimation methodology based on linear programming for measuring relative efficiencies of a collection of firms or entities (called Decision Making Units or DMUs) in transforming their inputs into outputs. A DEA domain is completely specified by a finite list of data points, one for each DMU. Each data point is a vector with components partitioned into two categories: those associated with "inputs" and those associated with "outputs" representing the activity of the DMU. The combined data about the DMUs and the assumptions about the technology define a *production possibility set*, the set of all points such that their relation among input and output components are theoretically feasible according to specified production criteria specific for the model.

We will work with the four production possibility sets which correspond to the four standard models in DEA. They are *i)* the constant returns model also known as the CCR model, introduced by Charnes, Cooper, and Rhodes [1978], *ii)* the variable returns, "BCC," model of Banker, Charnes, and Cooper [1984], and *iii)* & *iv)*, the increasing (IRS) and decreasing (DRS) returns models of Färe, Grosskopf, and Lovell [1985], and Seiford and Thrall [1990]. The reader is referred to these publications as well as to Banker and Thrall [1992] for a description, applicability, and explanation of the underlying assumptions for each of the four models.

A production possibility set is a convex polyhedral set finitely generated by the data domain. One objective of DEA is to "score" the units in the data domain. This consists of identifying the position of the DMU's corresponding data point in the production possibility set relative to a subset of the boundary points known as the *efficient frontier*. A data point which is on the efficient frontier corresponds to an "efficient" DMU; otherwise it is "inefficient." Another important aspect of an analysis can be the identification of reference points on the boundary of the production possibility set to be proposed as "benchmarks" for a given inefficient unit. Any DMU with input and output components which place it in the strict interior of the polyhedral set is inefficient as is any boundary point for which it is possible to decrease an input component or increase an output component and still remain in the set. All this is consistent with Koopman's [1951] treatment of efficiency.

A study involving any of the four DEA models may or may not be "oriented." If oriented, the analysis yields benchmark references for an inefficient DMU obtained by either decreasing inputs ("input" orientation) or increasing outputs ("output" orientation). If orientation-free, the analysis and results are based on the reduction of excesses; i.e., the slack variables. This is the case of the "additive" models (Charnes *et al.* [1985]) and their generalizations by Thrall [1996b]. Recent work offers yet more variations on orientation-free analysis; in particular, Cooper, Park, and Pastor [1998]. It is important to note that all these DEA forms, twelve in all, rely on the same four production possibility sets and their efficient subsets.

Data envelopment analysis is computationally intensive. The traditional process begins with a decision as to which of the four models will be employed and whether it will have input, output or no orientation. The procedure entails the solution of one linear program for each DMU in the un-oriented cases and, at least one and frequently two LPs in the oriented cases for each DMU (Arnold, *et al.* [1998]). Each DMU generates a slightly different LP. The size of the coefficient matrix of the LPs is (roughly) the number of inputs plus outputs times the number of DMUs.

Many published works address the problem of reducing computational times in DEA. One is to enhance the traditional approach. Several ideas have been proposed which do, in fact, have a significant impact on computational times (see, e.g., Ali [1993]). A recent contribution by Barr and Durchholz [1997] is based on partitioning the data set. They apply the principle that if a DMU is

inefficient with respect to a subset of the data points, it will also be inefficient with respect to the entire data set. Working with small subset means working with smaller LPs.

A fundamentally different procedure for DEA appears in Dulá [1998]. This approach begins with an efficient identification of a *frame* of a production possibility set; that is, a minimal set of data points needed to define the production possibility set and consequently sufficient to perform a complete DEA analysis on the full data set. A frame is specific to the DEA model; that is, a DEA data domain has frames for each of the models. The frame approach is favored when the problems are large with relatively few inputs and outputs and low efficiency density. The DMUs are scored in a second phase at which point orientation decisions are made.

In this paper we develop and test a new computational procedure involving frames of production possibility sets. The approach is based on an efficient identification of all four frames, one for each of the production possibility sets. To accomplish this we utilize relations among the four frames. Access to the frames permits expeditious scoring of the rest of the DMUs especially if the frame is a small subset of the data.

The first section of the paper presents the data, notation, assumptions and formal expression of the concepts with which we deal. It is here where we present the rigorous geometric definition of the polyhedral sets in the four DEA models and we introduce the definitions of the frames of the data domain. Section 2 formalizes the requisite relevant geometric results related to extreme elements in the four production possibility sets. Our results on how the four frames are interrelated are substantially strengthened extensions and formalizations of earlier works by Seiford and Thrall [1990], and more recently by Oppa and Yue [1997]. Section 3 presents a procedure, **All-Frames**, which exploits the interrelation among frames. The report on the computational testing of **AllFrames** is in Section 4. The paper closes with the conclusion that the new approach to DEA analysis based on the relations among the four frames offers significant increases in flexibility and speed. The paper has three appendices. The first is where all proofs have been relegated. The second is dedicated to a discussion on the impact of proportional points in the data domain. These are points that are multiples of each other. The last appendix contains a discussion of some benefits of using procedures based on frames specifically with regards to issues of weak efficiency.

1. The data domain and its four production possibility sets. The data domain for a DEA study is the set $\mathcal{A}$ of the n data points, $a^1, \ldots, a^n$; one for each DMU. Each data point is composed of two types of components, those pertaining to the m_1 inputs, $0 \neq x^j \geq 0$, and those corresponding to the m_2 outputs, $0 \neq y^j \geq 0$. We organize the data in the following way:

$$A = \left[a^1, \cdots, a^n \right], \quad \text{where,} \quad a^j = \begin{bmatrix} -x^j \\ y^j \end{bmatrix}; \quad j = 1, \ldots, n,$$

and A is the $m_1 + m_2$ by n matrix the columns of which are the data points.

We assume no proportionality among points; that is, no two points in the data domain are multiples of each other. Notice that this assumption also excludes duplicate data points. Proportionality is rare in real world data where measurements are independently made and only in special circumstances do proportional points have an impact on our results. However, proportionality is an important theoretical issue. Proportionality and duplication play different roles in different models. We have relegated our discussion of the impact of proportionality among data points in our procedures to an appendix.

The four polyhedral sets, or "hulls," are explicitly defined in what follows (Banker *et al.* [1984], Seiford and Thrall[1990], and Dulá and Venugopal [1995]):[‡]

$$\mathcal{P}^\ell = \left\{ z \in \Re^m \,\middle|\, z \leq A\lambda; \quad \text{for } \lambda \in \Lambda^\ell \right\}, \, \ell = 1, 2, 3, 4; \tag{1.1}$$

where Λ^ℓ is the subset of $\Lambda = \{\lambda | \lambda_j \geq 0, \, j = 1, \ldots, n\}$ defined by

$$
\begin{aligned}
CCR: && \Lambda^1 &= \Lambda; & (1.2a) \\
BCC: && \Lambda^2 &= \left\{ \lambda \in \Lambda \,\middle|\, \textstyle\sum_{j=1}^n \lambda_j = 1 \right\}; & (1.2b) \\
IRS: && \Lambda^3 &= \left\{ \lambda \in \Lambda \,\middle|\, \textstyle\sum_{j=1}^n \lambda_j \leq 1 \right\}; & (1.2c) \\
DRS: && \Lambda^4 &= \left\{ \lambda \in \Lambda \,\middle|\, \textstyle\sum_{j=1}^n \lambda_j \geq 1 \right\}. & (1.2d)
\end{aligned}
$$

[‡] Note, these representations for the four production possibility sets admit points with some zero and negative output values. See Pastor [1996]. If we wish to avoid negative values simply replace $\Re^m$ with $\Re^m_+$ in (1.1).

It is important to stress here that the production possibility sets are independent of orientation considerations. This implies that orientation does not affect the location of the production possibility set's efficient frontier, the subset of the boundary of $\mathcal{P}^\ell$ where the efficient DMUs reside. We will use $\mathcal{E}^\ell$ to denote the efficient frontier of $\mathcal{P}^\ell$.

Our development relies on the concept of minimal generating subsets of the data domain. Formally,

DEFINITION. A subset $\mathcal{G}$ of $\mathcal{A}$ is said to *generate* the production possibility set, $\mathcal{P}^\ell$, if

$$\mathcal{P}^\ell = \left\{ z \mid z \leq A\lambda \ \text{ for } \ \lambda \in \Lambda^\ell \ \text{ and } \ \lambda_j = 0 \ \text{ if } \ a^j \notin \mathcal{G} \right\}. \tag{1.3}$$

A minimal cardinality generating subset, $\mathcal{F}^\ell$, of $\mathcal{A}$ for the production possibility set $\mathcal{P}^\ell$ is called a *frame*. Our definition of frame means that if $\mathcal{F}^\ell$ is a frame and $a^{j^*} \in \mathcal{F}^\ell$ then the set of points $\mathcal{F}^\ell \backslash \{a^{j^*}\}$ cannot be a generating set for the production possibility set $\mathcal{P}^\ell$. Denote frames of the production possibility sets by $\mathcal{F}^1$, $\mathcal{F}^2$, $\mathcal{F}^3$, and $\mathcal{F}^4$ according to our conventions. We will prove $\mathcal{F}^2$, $\mathcal{F}^3$, and $\mathcal{F}^4$ are uniquely defined and show how to designate a unique frame for $\mathcal{P}^1$.

Any DEA inference about a DMU can be achieved using only the points in a frame for the corresponding production possibility set. Since the production possibility sets generated by the frame and by the entire data domain are the same, all questions concerning the efficiency status of a DMU can be answered using only the frame. Therefore, when the frame is a small subset of the data set, the DMUs can be processed using much smaller LPs. It is also important to note that our explicit (temporary) exclusion of proportional points in $\mathcal{A}$ implies there is a unique frame for each model. This is a consequence, in part, of the fact that the production possibilities sets always contain at least one extreme point obviating the possibility of lineality spaces (see Rockafellar [1970], Cor. 18.5.3 and Dulá *et al.* [1998])

2. Extreme-efficient DMUs, frames, and their interrelations. In this section we focus on a special category of efficient DMUs known as "extreme-efficient" since these relate directly to the frame of the data set. The definition of extreme-efficient DMUs first appeared in the paper by Charnes, Cooper, and Thrall in [1986] and was discussed in more depth in the sequel [1991].

The concept of extreme-efficiency in these two papers applies to the constant returns model and is not connected as directly as we require it with the geometry of the production possibility set. The objective of this section is to generalize the concept of extreme efficiency to the other three models and to connect it directly to the geometry of production possibility sets and the concept of its frame. This section is also where we establish the connections among the four frames.

The production possibility sets corresponding to the four DEA models are polyhedral sets defined by constrained linear combinations of the data along with a set of directions of recession. The set $\mathcal{P}^1$ is the only one which is a cone since if $\hat{z} \in \mathcal{P}^1$ so does $\alpha\hat{z}$ for any $\alpha \geq 0$. The efficient frontier, $\mathcal{E}^\ell$, of $\mathcal{P}^\ell$ is defined by extreme elements which are extreme points in the case of the variable, increasing, and decreasing returns models and extreme rays in the constant returns model. These essential extreme elements correspond to DEA data points and a frame is a set of data points that are extreme elements of $\mathcal{E}^\ell$ (see Dulá et al. [1998]). We will establish the relation between points in the frame and the extreme-efficient DMUs as defined next.

DEFINITION. For the four DEA models, $\ell = 1, 2, 3$, or 4, DMU j^* is extreme-efficient if and only if the system

$$A\lambda \geq a^{j^*}, \ \lambda \in \Lambda^\ell, \ \lambda_{j^\cdot} = 0; \tag{2.1}$$

is infeasible.

This definition is a natural extension of the conditions for extreme-efficiency in the case of the constant returns model given by Charnes, Cooper, and Thrall in [1991] (see Theorem 7C.(E4), p. 215). Note too that this definition for extreme-efficiency is orientation-free; that is, it is independent of whether or not the analysis is oriented. Finally, it should be clear that if a DMU satisfies a model's condition for extreme-efficiency with respect to the full data domain $\mathcal{A}$, it satisfies it with respect to any generating subset – including the frame.

The following result establishes the correspondence between extreme-efficiency and the corresponding frame of the data. This theorem will be useful in our proofs for the two results which follow it.

THEOREM 1. *For each of the four DEA models, a point belongs to the model's frame if and only if it is extreme-efficient in that model.*

PROOF. See Appendix A.

Next we present the results which establish the properties and relations between and among the four frames we employ in the design of a new computational framework for DEA.

THEOREM 2. $\mathcal{F}^2 = (\mathcal{F}^3 \cup \mathcal{F}^4)$.

PROOF. See Appendix A.

THEOREM 3. $\mathcal{F}^1 = (\mathcal{F}^3 \cap \mathcal{F}^4)$.

PROOF. See Appendix A.

COROLLARY.

   a. $\mathcal{F}^1 \subset \mathcal{F}^3 \subset \mathcal{F}^2$.

   b. $\mathcal{F}^1 \subset \mathcal{F}^4 \subset \mathcal{F}^2$.

PROOF. Direct consequence of set operations on the two previous results.

These three results establish that the frame for the production possibility set of the variable returns model contains the other three frames. Theorem 3 has immediate computational implications since it essentially states that after calculating any two frames from the list, $\{\mathcal{F}^3, \mathcal{F}^4, \mathcal{F}^1\}$, the third may be obtained through simple set operations. This suggests several ideas for procedures to score DMUs in all four models and orientations based on identifying $\mathcal{F}^2$ first and from there extracting two more frames directly and inferring the last by simple set operations. One of these ideas is tested in the next section.

3. Procedure AllFrames. We propose to exploit the relations among frames found in the previous section to develop a procedure, **AllFrames**, to identify all four frames of a given DEA

data set. Consider three routines VRFRAME, IRFRAME, and DRFRAME the inputs of which are DEA data sets and outputs will be frames as follows:

$$\mathcal{F}^2 \leftarrow \text{VRFRAME}(\mathcal{A}) \tag{3.1a}$$

$$\mathcal{F}^3 \leftarrow \text{IRFRAME}(\mathcal{A}) \tag{3.1b}$$

$$\mathcal{F}^4 \leftarrow \text{DRFRAME}(\mathcal{A}) \tag{3.1c}$$

Routines for identifying the frame of the data for the variable, increasing, and decreasing returns models could be based on results in Dulá and Hickman [1997]. This work presents necessary and sufficient conditions for a DMU to be extreme-efficient based on the solution of any form (oriented or otherwise) of the envelopment LPs by deleting the DMU being scored from the input-output matrix. A specialized routine for identifying frames in DEA appears in Dulá [1998]. The approach used in Dulá [1998] "builds" the frame one element at a time by exploiting the geometrical properties of the production possibility sets in DEA. The algorithms are an extension of work by Dulá and others on frames for general polyhedral sets (see Dulá and Helgason [1996] and Dulá *et al.* [1998]).

An important realization is that routines IRFRAME and DRFRAME can be used as follows:

$$\mathcal{F}^3 \leftarrow \text{IRFRAME}(\mathcal{F}^2) \tag{3.2a}$$

$$\mathcal{F}^4 \leftarrow \text{DRFRAME}(\mathcal{F}^2) \tag{3.2b}$$

This follows from the fact that $\mathcal{F}^3 \subset \mathcal{F}^2 \subset \mathcal{A}$ and $\mathcal{F}^4 \subset \mathcal{F}^2 \subset \mathcal{A}$. With this we have enough to design the new procedure for DEA analysis based on the identification of frames.

PROCEDURE AllFrames

PHASE 1. Identify all four frames of the DEA data set $\mathcal{A}$.

    <u>Step 1.</u> $\mathcal{F}^2 \leftarrow \text{VRFRAME}(\mathcal{A})$.

    <u>Step 2.</u> $\mathcal{F}^3 \leftarrow \text{IRFRAME}(\mathcal{F}^2)$.

    <u>Step 3.</u> $\mathcal{F}^4 \leftarrow \text{DRFRAME}(\mathcal{F}^2)$.

Step 4. $\mathcal{F}^1 \leftarrow \mathcal{F}^3 \cap \mathcal{F}^4$.

PHASE 2. Select model and orientation and score DMUs using appropriate frame.

END PROCEDURE AllFrames.

Let us analyze some scenarios. In the worst case, the cardinalities of the frame and the entire data set are almost the same. In this case, the effort to find the the frames of the data for the four models will require roughly three times the time to find the frame of the variable returns model. This since step four is computationally inexpensive. So, in a sense, we obtain four frames for the price of three. In real applications, however, such "dense" data domains do not seem to be the norm. A more realistic scenario is that the set of efficient DMUs is relatively small compared to the full DEA data set. This is especially true for large problems. Reports indicate that this relation can be less than 1% (see, e.g. Barr and Durchholz [1997] where they report DEA efficiencies of a study on nearly 9000 U.S. banks using Federal Reserve data where fewer than 87 are efficient). We may expect then to be able to complete the four steps in Phase 1 in little more than the time needed to execute Step 1 alone. This expectation is actually realized as we will see in Section 4, below.

The frame of a DEA data set depends on the model and not on any orientation consideration. This translates into more flexibility since orientation selections are postponed to the beginning of Phase 2 when much of the heavy work is behind us. This means that "scoring" can proceed picking and choosing any desired combinations of model and orientations with the frames. With the frame in hand, the second phase to score the DMUs can proceed using LPs with dimensions defined by the frame instead of the entire DEA data set. If the frames are relatively small computation time for scoring is substantially reduced.

4. Computational results. In order to test and validate **AllFrames**, the new procedure was coded using the frame routines in Dulá [1998] and applied to 40 DEA problems.

The 40 problems in our suite were randomly generated to reflect the types of situations that occur in DEA. The two parameters of the problem generator are the dimension (inputs plus outputs) and number of DMUs. It was considered for our suite that problems should have fewer than twenty

dimensions. However, there is great variability in the number of DMUs which may occur in practice. Problems become computationally interesting when they have more than 100 DMUs. Considering that problems with almost 9000 DMUs are being formulated and solved, DEA problems were randomly generated for 5, 10, and 20 dimensions and 125, 250, 500, 1000, 2500, 5000, and 10000 DMUs. Another problem characteristic is the ratio of efficient DMUs to the total number of points; what we call the *efficiency density* of the problem. It is difficult to control efficiency density of randomly generated problems; therefore, we measured it *expo facto*, making it itself random. Note that higher efficiency densities become increasingly improbable for problems with few dimensions when the number of DMUs is large. For this reason it was not possible to offer the full range of efficiency densities in all dimensions. All the DEA data files are available for examination at the author's web site.[†]
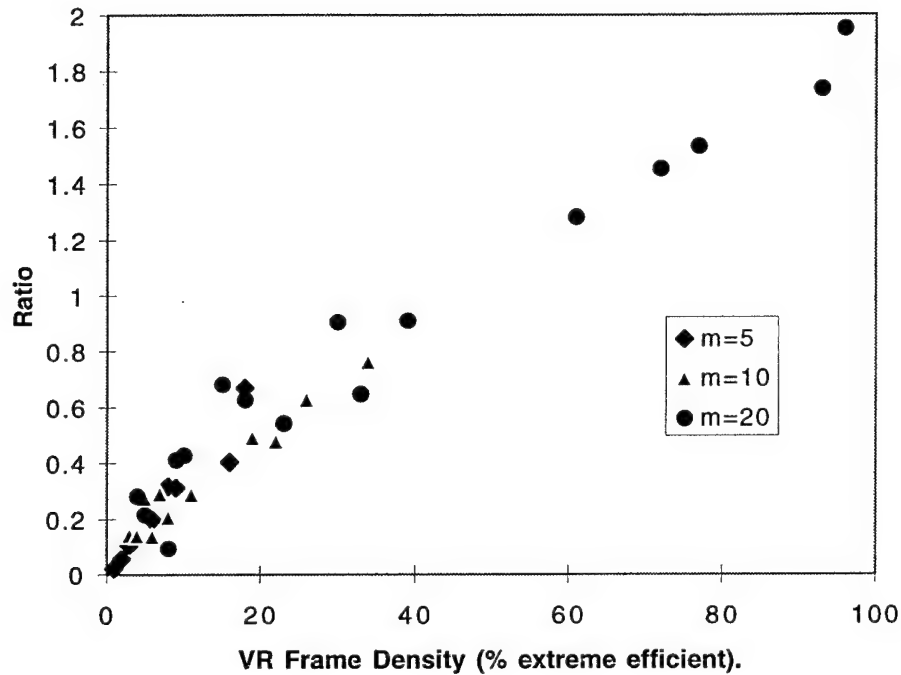
Table 1 collects the relevant results of our tests. The first column of the table contains the problem name. The first two digits in a problem name are the dimension. The following four digits correspond to the number of DMUs where 0000 is 10,000 and the last two give the efficiency density. The next four columns are the cardinalities of the frames for the variable, increasing, decreasing, and constant returns model for the corresponding DEA data file. The next four columns contain different execution times. Times are CPU plus system in seconds exclusive of reading data files and writing output. The times are virtually independent of the system load because the system I/O times are a small fraction of the total times; frequently less than 1%. Also, times generally did not vary significantly over several repetitions making it unnecessary to record more than one reading. All tests were serially performed using an SGI Power Challenge L with four R8000 processors at 75 Mhz.

The entries in column labeled "Traditional/(Additive)" correspond to the execution times using the standard approach in DEA based on solving the envelopment LP for every DMU. This was done for the orientation-free "additive" form of the variable returns model.[‡] The entries in the "VR"

[†] http://www.olemiss.edu/~jdula/DEADATA

[‡] The standard additive model uses the sum of the slacks as the objective function. Although this has been criticized (see e.g. Thrall [1996a, 1996b]) it is still an appropriate representative for our present computational purposes and in particular for checking DEA efficiency.

Figure 1. Performance ratio TIME(IR+DR+CR)/TIME(VR).

---

column are the times to find the frame of the variable returns model using the frame algorithm in Dulá [1998]. Dulá's frame procedure was applied again to the variable returns frame to extract the increasing and decreasing returns frames. The constant returns frame was identified by finding the simple intersection of the last two frames. The times to find these three frames are recorded in the column labeled "IR + DR + CR." The second phase to score the rest of the DMUs can proceed once access to the frame of a model is available. This was done for the variable returns "additive" model and the times recorded in column "Phase 2/(Additive)." This serves as comparison with the traditional approach. The last column contains the ratio of the times to find the frames of the increasing, decreasing, and constant returns models with the time to find the frame of the variable returns model.

The essence of the data is best captured and understood by the plot in Figure 1. This plot depicts the relation between the data in the last column of Table 1 and the variable returns frame density (column "VR" divided by the number of DMUs). From the chart we confirm what we anticipated earlier about procedure **AllFrames**; namely, that the performance in Phase 1 is directly proportional to the density of the frame of the variable returns model. This means that

procedure **AllFrames** provides substantial time benefits when the frame density is low and is less favored at high densities. At high densities, the prediction that the four frames are found in less than three times the time to find the variable returns frame is realized. In practice, though, the percent of extreme-efficient DMUs tends to be low as in the case of DEA file "05by0000at01." The point in the chart corresponding to this file is the lowest on the left. This indicates that for a bit more time to find one frame we have access to all four frames. This is the ideal situation for procedure **AllFrames**. With such problems we can count on substantial time savings on DEA studies involving multiple models and orientations.

5. Concluding remarks. The special relation among the four frames of the standard DEA models along with the fact that these frames tend to be small subsets of the DEA data set offers the possibility of reducing times while increasing flexibility of DEA studies especially when these are over several models and multiple orientations. These interactions among frames can be exploited in several ways based essentially on the fact that the variables returns frame is the superset for all frames and that two of the remaining three frames are sufficient to acquire the fourth one almost free. One such procedure, **AllFrames**, was actually implemented based on extracting the increasing and decreasing returns frames from the variable returns frame and obtaining the constant returns frame through simple set operations. The procedure was coded to validate the concept and measure its performance. The results confirm that, without much additional effort than what is required for a single study with a fixed model and orientation, the study can be extended to all four models and several choices of orientations. Access to procedures such as **AllFrames** will encourage exploration in DEA studies leading to a more efficient and effective application of DEA. The results in this paper will be the basis for future work on reducing computational times for DEA analyses over multiple models and orientations. One focus will be on investigating different sequences for finding the frames. Also promising is to explore using information about supports of the production possibility set of the variables returns model and using this in conjunction with the constant returns frame to extract the remaining two frames.

Table 1.
Implementation Results of **AllFrames**

DEA File	Cardinalities of the Frames				Times*				Ratio
	VR	IR	DR	CR	Traditional (Additive)	VR	Phase 1. IR+DR+CR*	Phase 2. (Additive)	$\frac{IR+DR+CR}{VR}$
05by0125at09	11	10	5	4	0.30	0.11	0.03	0.08	0.27
05by0250at09	22	21	20	19	1.24	0.37	0.11	0.25	0.30
05by0500at08	38	35	35	32	6.43	0.81	0.26	0.97	0.32
05by1000at06	63	58	37	32	24.89	2.28	0.45	2.55	0.20
05by2500at03	84	74	49	39	139.05	6.11	0.64	7.26	0.11
05by5000at02	84	74	49	39	545.71	12.92	0.72	14.18	0.06
05by0000at01	63	58	37	32	1414.91	21.26	0.43	15.45	0.02
05by0125at18	22	21	20	19	0.40	0.16	0.11	0.13	0.67
05by0250at16	39	38	15	14	1.75	0.56	0.22	0.42	0.40
10by0125at06	8	8	8	8	0.47	0.25	0.03	0.14	0.14
10by0250at08	20	19	18	17	2.12	0.57	0.12	.046	0.21
10by0500at07	36	36	30	30	5.34	1.04	0.30	0.58	0.29
10by1000at05	54	54	8	8	22.73	1.16	0.32	1.79	0.27
10by2500at04	101	101	101	17	130.66	6.77	0.95	6.59	0.14
10by5000at05	264	232	232	159	651.41	55.25	10.48	34.01	0.19
10by0000at03	264	232	232	159	2185.26	63.93	9.09	59.36	0.14
10by0125at19	24	24	15	15	0.53	0.28	0.14	0.15	0.49
10by0250at22	54	54	11	11	2.46	0.90	0.43	0.62	0.48
10by0500at34	171	163	117	109	14.11	5.80	4.42	3.64	0.76
10by1000at26	264	232	191	159	55.03	14.49	9.09	11.35	0.63
10by2500at11	264	232	191	159	222.64	36.35	10.49	21.44	0.29
20by0125at08	10	10	10	10	0.99	0.70	0.07	0.36	0.09
20by0250at10	25	25	25	25	3.90	1.65	0.30	1.01	0.18
20by0500at10	49	49	49	49	14.82	2.62	0.83	2.59	0.32
20by1000at10	96	96	95	95	36.69	5.68	2.43	3.45	0.43
20by2500at09	226	224	219	217	262.59	28.61	11.76	21.51	0.41
20by5000at05	226	224	219	217	958.33	54.11	11.58	44.19	0.21
20by0000at04	386	375	379	368	3930.54	216.44	61.13	146.58	0.28
20by0125at39	49	49	49	49	1.09	1.04	0.95	0.28	0.91
20by0250at23	58	58	38	38	3.20	1.40	0.76	0.80	0.54
20by0500at33	163	162	63	62	15.14	7.16	4.64	3.47	0.65
20by1000at18	181	181	167	167	57.18	13.71	8.61	7.41	0.63
20by2500at15	386	375	379	368	345.25	60.72	60.72	37.15	0.68
20by5000at30	1521	1473	1437	1389	2557.81	1000.97	907.79	430.76	0.91
20by0000at15	1521	1473	1437	1389	7510.29	1343.99	913.24	770.30	0.68
20by0125at96	120	118	120	118	1.83	3.16	6.17	0.109	1.95
20by0250at93	232	231	227	226	7.03	11.55	20.05	0.59	1.74
20by0500at77	386	375	379	368	29.00	33.75	51.76	5.88	1.53
20by1000at72	717	710	686	679	121.03	120.97	175.80	26.27	1.45
20by2500at61	1521	1473	1437	1389	823.66	649.98	832.65	213.89	1.28

*Times in seconds
*Calculations using the VR frame.

References.

Ali, I., [1993], "Streamlined computation for data envelopment analysis," *European Journal of Operational Research,* Vol. 64, pp. 61-67.

Arnold, V., I. Bardham, W.W. Cooper, and A. Gallegos, [1998],"Primal and dual optimality in computer codes using two-stage solution procedures in DEA," in *Operations Research: Methods, Models, and Applications,* J.E. Aronson and S. Zionts, eds., Quorum Books, Westport, CT.

Banker, R.D., A. Charnes, and W.W. Cooper, [1984], "Some models for estimating technological and scale inefficiencies in data envelopment analysis," *Management Science,* Vol. 30, No. 9, pp. 1078–1092.

Banker, R.D., and R.M. Thrall, [1992], "Estimation of returns to scale using Data Envelopment Analysis," *European Journal of Operational Research,* Vol. 62, pp. 74–84.

Barr, R.S. and M.L. Durchholz, [1997], "Parallel and hierarchical decomposition approaches for solving large-scale Data Envelopment Analysis models," *Annals of Operations Research,* 73, pp. 339–372.

Charnes, A., W.W. Cooper, and E. Rhodes, [1978], "Measuring the efficiency of decision making units," *European Journal of Operational Research,* Vol. 2, No. 6, pp. 429-444.

Charnes, A., W.W. Cooper, B. Golany, L. Seiford, and J. Stutz, [1985], "Foundations of data envelopment analysis for Pareto-Koopmans efficient empirical production functions," *Journal of Econometrics,* Vol. 30, pp. 91–107.

Charnes, A., W.W. Cooper, and R.M. Thrall, [1986], "Identifying and classifying scale and technical efficiency in Data Envelopment Analysis," *Operations Research Letters,* Vol. 5, pp. 105–116.

Charnes, A., W.W. Cooper, and R.M. Thrall, [1991], "A structure for classifying and characterizing efficiency and inefficiency in data envelopment analysis," *The Journal of Productivity Analysis,* Vol. 2, pp. 197–237.

Cooper, W.W., K.S. Park, and J.T. Pastor, [1998], "RAM: A range adjusted measure of inefficiency for use with additive models and relations to other models and measures in DEA," Working Paper, IC^2 Institute, The University of Texas, Austin, TX 78712-1174. Accepted in *Journal of Productivity Analysis.*

Dulá, J.H., [1998], "An algorithm for Data Envelopment Analysis (DEA)." Submitted to *INFORMS Journal on Computing.*

Dulá, J.H. and N. Venugopal, [1995], "On characterizing the production possibility set for the CCR ratio model in DEA," *International Journal of Systems Science* Vol. 26, No. 12, pp. 2319-2325.

Dulá, J.H. and R.V. Helgason, [1996], "A new procedure for identifying the frame of the convex hull of a finite collection of points in multidimensional space," *European Journal of Operational Research,* Vol. 92, pp. 352–367.

Dulá, J.H., R.V. Helgason, and N. Venugopal [1998], "An algorithm for identifying the frame of a pointed finite conical hull," *INFORMS Journal on Computing*, Vol. 10, No. 3.

Dulá, J.H. and B. L. Hickman, [1997], "Effects of excluding the column being scored from the DEA envelopment LP technology matrix," *Journal of the Operational Research Society*, Vol. 48, pp. 1001–1012.

Färe, R., S. Grosskopf, C.A.K. Lovell, [1985], *The Measurement of Efficiency Production.* Kluwer Academic Publishers, Boston.

Oppa, G. and M. Yue [1997], "On measuring scale efficiency in DEA," unpublished manuscript, London School of Economics, London, UK.

Pastor, J., [1996], "Translation invariance in Data Envelopment Analysis: A generalization," *Annals of Operations Research*, Vol. 66, pp. 93–102.

Koopman, T., [1951], *Activity Analysis of Production and Allocation*, John Wiley & Sons, Inc., New York.

Rockafellar, R.T., [1970], *Convex Analysis*, Princeton University Press, Princeton, NJ.

Seiford, L.M., and R.M. Thrall, [1990], "Recent developments in DEA; the mathematical programming approach to frontier analysis," *Journal of Econometrics*, Vol. 46, pp. 7–38.

Thrall, R.M., [1996a], "Duality, classification and slacks in DEA," *Annals of Operations Research*, Vol. 66, pp. 109–138.

Thrall, R.M., [1996b], "Goal vectors for DEA efficiency and inefficiency," Working Paper No. 128, J.H. Jones Graduate School of Administration, Rice University, Houston, TX 77005-1892.

Appendix A. Proofs

THEOREM 1. *For each of the four DEA models, a point belongs to the model's frame if and only if it is extreme-efficient in that model.*

PROOF. The proof applies for all four DEA models. Recall that our assumptions exclude proportional points in the data domain $\mathcal{A}$. However, it must be preceded by the following general lemma establishing an important property of the frame, $\mathcal{F}^\ell$, for the data set $\mathcal{A}$ in any of the four DEA models: $\ell; \ell = 1, 2, 3,$ or 4 as per our conventions.

LEMMA 1. *If $a^{j^*} \in \mathcal{F}^\ell$ then the system*

$$\sum_{\{j|a^j \in \mathcal{F}^\ell\}-j^*} a^j \lambda_j \geq a^{j^*}, \quad \lambda \in \Lambda^\ell; \tag{A.1}$$

has no solution.

Proof of Lemma 1. Suppose $\hat{\lambda} \in \Lambda^\ell$ solves System (A.1); that is,

$$\sum_{\{j|a^j \in \mathcal{F}^\ell\}-j^*} a^j \hat{\lambda}_j \geq a^{j^*}. \tag{A.2}$$

Now consider any point $\tilde{a} \in \mathcal{P}^\ell$. Then since $\mathcal{F}^\ell$ is a frame, there exists a solution, $\tilde{\lambda} \in \Re^n$; $\tilde{\lambda}_j = 0$ for $j \notin \{j|a^j \in \mathcal{F}^\ell\}$, and

$$\sum_{\{j|a^j \in \mathcal{F}^\ell\}} a^j \tilde{\lambda}_j = \sum_{\{j|a^j \in \mathcal{F}^\ell\}-j^*} a^j \tilde{\lambda}_j + a^{j^*} \tilde{\lambda}_{j^*} \geq \tilde{a}. \tag{A.3}$$

By (A.2) we can substitute a^{j^*} in (A.3) without affecting the inequality as follows:

$$\tilde{a} \leq \sum_{\{j|a^j \in \mathcal{F}^\ell\}-j^*} a^j \tilde{\lambda}_j + a^{j^*} \tilde{\lambda}_{j^*} \leq \sum_{\{j|a^j \in \mathcal{F}^\ell\}-j^*} a^j \tilde{\lambda}_j + (\underbrace{\sum_{\{j|a^j \in \mathcal{F}^\ell\}-j^*} a^j \hat{\lambda}_j}_{\geq a^{j^*}, by\ (A.2).}) \tilde{\lambda}_{j^*}. \tag{A.4}$$

Therefore,

$$\sum_{\{j|a^j \in \mathcal{F}^\ell\}-j^*} a^j \tilde{\lambda}_j + (\sum_{\{j|a^j \in \mathcal{F}^\ell\}-j^*} a^j \hat{\lambda}_j) \tilde{\lambda}_{j^*} = \sum_{\{j|a^j \in \mathcal{F}^\ell\}-j^*} a^j (\tilde{\lambda}_j + \hat{\lambda}_j \tilde{\lambda}_{j^*}) \geq \tilde{a}. \tag{A.5}$$

Set

$$\lambda_j' = \begin{cases} \tilde{\lambda}_j + \hat{\lambda}_j \tilde{\lambda}_{j^*} & \text{if } j \in \{j | a^j \in \mathcal{F}^\ell\} - j^*; \\ 0, & \text{otherwise.} \end{cases} \tag{A.6}$$

This makes λ' a solution to (A.3). It remains to be shown that $\lambda' \in \Lambda^\ell$. This only needs to be done for the variable, increasing and decreasing returns cases since in all four cases nonnegativity is obvious; what needs to be verified is that the sum of the coefficients corresponds to the model's requirement

$$\sum_j \lambda_j' = \sum_j(\tilde{\lambda}_j + \hat{\lambda}_j\tilde{\lambda}_{j^*}) = \sum_j \tilde{\lambda}_j + \tilde{\lambda}_{j^*}\underbrace{\sum_j \hat{\lambda}_j}_{=1} = \sum_j \tilde{\lambda}_j + \tilde{\lambda}_{j^*} = 1; \quad \text{(VR case)}.$$

$$\sum_j \lambda_j' = \sum_j(\tilde{\lambda}_j + \hat{\lambda}_j\tilde{\lambda}_{j^*}) = \sum_j \tilde{\lambda}_j + \tilde{\lambda}_{j^*}\underbrace{\sum_j \hat{\lambda}_j}_{\leq 1} \leq \sum_j \tilde{\lambda}_j + \tilde{\lambda}_{j^*} \leq 1; \quad \text{(IR case)}.$$

$$\sum_j \lambda_j' = \sum_j(\tilde{\lambda}_j + \hat{\lambda}_j\tilde{\lambda}_{j^*}) = \sum_j \tilde{\lambda}_j + \tilde{\lambda}_{j^*}\underbrace{\sum_j \hat{\lambda}_j}_{\geq 1} \geq \sum_j \tilde{\lambda}_j + \tilde{\lambda}_{j^*} \geq 1; \quad \text{(DR case)}.$$

With this we may conclude that $\mathcal{F}^\ell$ is not minimal, and hence not a frame, since any arbitrary point in $\mathcal{P}^\ell$ can be represented without a^{j^*}. This contradiction establishes the proof. ∎

Proof of Theorem 1. By Lemma 1 above, $a^{j^*} \in \mathcal{F}^\ell$ means the system

$$\sum_{\{j | a^j \in \mathcal{F}^\ell\} - j^*} a^j \lambda_j \geq a^{j^*}; \tag{A.7}$$

is infeasible. This is sufficient to conclude that a^{j^*} is extreme efficient. For the reverse implication, if a point corresponds to an extreme-efficient DMU then the corresponding system (2.1) is infeasible. This means that $\mathcal{P}^\ell$ cannot be generated without a^{j^*}. Therefore, a^{j^*} must be in every generating set for $\mathcal{P}^\ell$ including, of course, the frame. ∎

THEOREM 2. $\mathcal{F}^2 = (\mathcal{F}^3 \cup \mathcal{F}^4)$.

Proof. $\underline{\mathcal{F}^2 \subset (\mathcal{F}^3 \cup \mathcal{F}^4)}$. $a^{j^*} \in \mathcal{F}^2$ means system (2.1) has no solution for $\ell = 2$. Then either there is no nonnegative solution to the system $\sum_{\substack{j=1 \\ j \neq j^*}}^{n} a^j \lambda_j \geq a^{j^*}$ or, if there is, none is such that $\sum_{\substack{j=1 \\ j \neq j^*}}^{n} \lambda_j = 1$. In the first case the result follows since $a^{j^*} \in \mathcal{F}^3$ and $a^{j^*} \in \mathcal{F}^4$. Suppose it was the second case. By showing that it is impossible to have two solutions such $\hat{\lambda} \in \Lambda^3$ and $\tilde{\lambda} \in \Lambda^4$, we will have shown that $a^{j^*} \in \mathcal{F}^3$ or $a^{j^*} \in \mathcal{F}^4$. Note that neither solution can have its components add up to unity and, by our assumptions on the data domain, $\hat{\lambda}, \tilde{\lambda} \neq 0$. Then, there exists a coefficient $0 < \gamma < 1$ such that

$$\gamma \sum_{\substack{j=1 \\ j \neq j^*}}^{n} \hat{\lambda}_j + (1 - \gamma) \sum_{\substack{j=1 \\ j \neq j^*}}^{n} \tilde{\lambda}_j = 1. \tag{A.8}$$

We can use this coefficient to construct a new solution, $\bar{\lambda} = \gamma \hat{\lambda} + (1 - \gamma) \tilde{\lambda} \geq 0$, which is also feasible but the components of which add up to unity contradicting our initial premise.

$\underline{(\mathcal{F}^3 \cup \mathcal{F}^4) \subset \mathcal{F}^2}$. Suppose $a^{j^*} \in (\mathcal{F}^3 \cup \mathcal{F}^4)$, then either $a^{j^*} \in \mathcal{F}^3$ or $a^{j^*} \in \mathcal{F}^4$. In either case, system (2.1) with either $\ell = 3$ or $\ell = 4$ has no solution which means that a solution is also impossible for the more restrictive case when $\sum_{\substack{j=1 \\ j \neq j^*}}^{n} \lambda_j = 1$ implying $a^{j^*} \in \mathcal{F}^2$. ∎

THEOREM 3. $\mathcal{F}^1 = (\mathcal{F}^3 \cap \mathcal{F}^4)$.

Proof. $\underline{(\mathcal{F}^3 \cap \mathcal{F}^4) \subset \mathcal{F}^1}$. $a^{j^*} \in (\mathcal{F}^3 \cap \mathcal{F}^4)$ means that the system (2.1) cannot have a solution for $\ell = 1$ since, otherwise, either $\sum_{\substack{j=1 \\ j \neq j^*}}^{n} \lambda_j \leq 1$ or $\sum_{\substack{j=1 \\ j \neq j^*}}^{n} \lambda_j \geq 1$ implying a^{j^*} either belongs to $\mathcal{F}^3$ or $\mathcal{F}^4$. The infeasibility of this system is sufficient to conclude $a^{j^*} \in \mathcal{F}^1$.

$\underline{\mathcal{F}^1 \subset (\mathcal{F}^3 \cap \mathcal{F}^4)}$. If $a^{j^*} \in \mathcal{F}^1$ and since, by assumption, there is no other point in $\mathcal{F}^2$ proportional to it, the system (2.1) with $\ell = 1$ has no solution. Since a more restrictive system cannot be feasible we can conclude that $a^{j^*} \in \mathcal{F}^3$ and $a^{j^*} \in \mathcal{F}^4$ and hence in the intersection. ∎
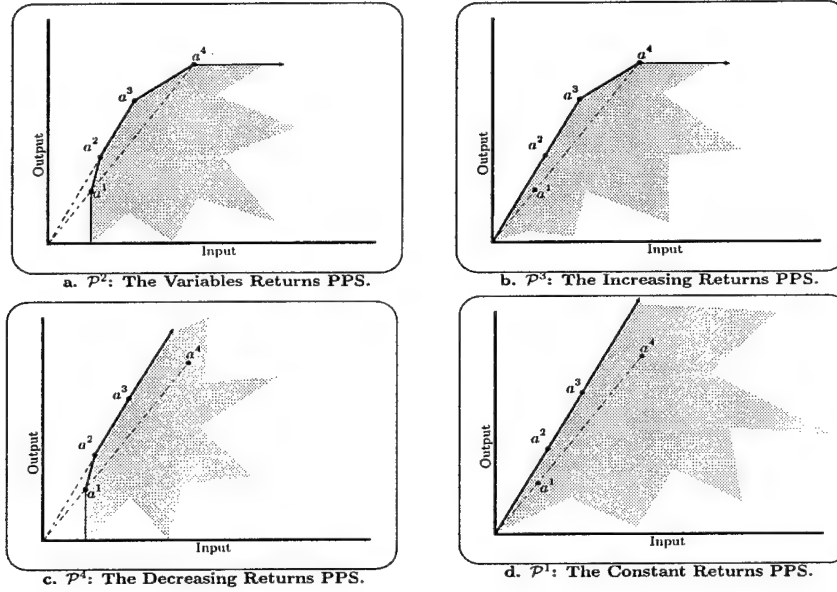
Appendix B. The case of proportionality in the data. We have relegated the discussion on the issue of proportionality among data points to this appendix because it is computationally simple to address but theoretically complex. We will discuss briefly some of the theoretical aspects before we proceed to presenting the modifications of procedure **AllFrames** needed to adapt it for the case when proportional points are not known to be absent.

The assumption of nonproportionality among data points in the data domain appears in previous published works (see e.g., Charnes *et al.* [1991], expression (8), p. 202 or Dulá and Hickman [1997]). This assumption obviates duplication of data points, another common assumption. In a work that deals with all four DEA models – constant, variable, increasing, and decreasing returns – simultaneously, such as the present work, these two assumptions play different roles. The weaker condition of no duplication of data points is necessary for all four cases for the frames to be unique. However, it is sufficient only for the variable, increasing, and decreasing returns models. Uniqueness of the frame is guaranteed for the constant returns model only if there is no proportionality.

Refer to Figure 2 to illustrate the impact of proportional points on the four DEA models and on our principal results. This figure depicts the four production possibility sets for the same data domain: $\mathcal{A} = \{a^1, a^2, a^3, a^4\}$. Notice that pairs of points $\{a^2, a^3\}$ and $\{a^1, a^4\}$ are proportional (the dashed lines serve to emphasize this relation). The frames for $\mathcal{P}^2$, $\mathcal{P}^3$, and $\mathcal{P}^4$ are $\mathcal{F}^2 = \{a^1, a^2, a^3, a^4\}$, $\mathcal{F}^3 = \{a^3, a^4\}$, and $\mathcal{F}^4 = \{a^1, a^2\}$, respectively, and these are unique. The first important realization is that $\mathcal{P}^1$ has two frames: $\hat{\mathcal{F}}^1 = \{a^2\}$ and $\tilde{\mathcal{F}}^1 = \{a^3\}$. This loss of uniqueness is a consequence of the fact that the points a^2 and a^3 are a proportional pair; that is, $a^2 = \alpha a^3$ where $0 < \alpha \neq 1$. This causes the invalidation of Theorems 1 and 3 and the Corollary in Section 2. With proportionality, Theorem 1 is no longer true for the constant returns model and Theorem 3 must be made weaker; i.e.: $\mathcal{F}^1 \subset (\mathcal{F}^3 \cap \mathcal{F}^4)$.

Fortunately for computations, proportional points behave predictably in their relation to frames and can be handled efficiently. One reason for this is the fact that if there is a set of two or more proportional points, only the one with the smallest and the one with the largest norm can belong to any of the frames $\mathcal{F}^2, \mathcal{F}^3, \mathcal{F}^4$ (anything in between is necessarily not extreme-efficient). If a proportional pair belongs to any of these frames at all, then: *i)* both points in the pair must belong to $\mathcal{F}^2$; *ii)* the longest of the two necessarily belongs to $\mathcal{F}^3$ but not to $\mathcal{F}^4$; and *iii)* the shortest

---

Figure 2.
The four production possibility sets from the same data set with proportionality.

a. $\mathcal{P}^2$: The Variables Returns PPS.

b. $\mathcal{P}^3$: The Increasing Returns PPS.

c. $\mathcal{P}^4$: The Decreasing Returns PPS.

d. $\mathcal{P}^1$: The Constant Returns PPS.

---

of the two necessarily belongs to $\mathcal{F}^4$ but not to $\mathcal{F}^3$. This means that the issue of proportionality becomes relevant only after the three frames, $\mathcal{F}^2, \mathcal{F}^3$, and $\mathcal{F}^4$ have been found.

Only proportional pairs in $\mathcal{F}^2$ which are also efficient with respect to the constant returns model are eligible to belong in $\mathcal{F}^1$. This is illustrated in Figure 2 with the proportional pair $\{a^2, a^3\}$ being an eligible pair while $\{a^1, a^4\}$ is not. Exactly one point from each eligible pair ends up in $\mathcal{F}^1$. The complication with proportional points is a consequence of the fact that neither from an eligible pair appears in $\mathcal{F}^3 \cap \mathcal{F}^4$. The example connected with Figure 2 can be used to verify this.

Therefore we need to identify eligible proportional pairs in $\mathcal{F}^2$ and provide an unambiguous rule for selecting which one from such a pair is to appear in $\mathcal{F}^1$. A rule to distinguish between eligible and ineligible proportional pairs in $\mathcal{F}^2$ is a consequence of the following lemma:

LEMMA. *Let a^{j_1} and a^{j_2} be two points in $\mathcal{A}$ such that $a^{j_2} = \alpha a^{j_1}$ for $\alpha > 1$ and suppose $a^0 = \gamma a^{j_1} + (1-\gamma)a^{j_2} \in \mathcal{E}^\ell$ for some $0 < \gamma < 1$. Then a^{j_1}, a^{j_2}, and a^0 are all efficient with respect to the constant returns DEA model.*

Proof. Since a^0 is efficient with respect to the variable returns model, there exist $0 < \pi^* \in \Re^m$, and $\beta^* \in \Re$ such that

$$\langle \pi^*, a^0 \rangle + \beta^* = 0, \qquad\qquad (B.1.1)$$

$$\langle \pi^*, a^j \rangle + \beta^* \leq 0; \qquad a^j \in \mathcal{A}. \qquad\qquad (B.1.2)$$

Notice that (B.1.1) implies $\langle \pi^*, a^{j_1} \rangle + \beta^* = \langle \pi^*, a^{j_2} \rangle + \beta^* = 0$ since otherwise a^0 could not be expressed as the strict convex combination of a^{j_1} and a^{j_2}. Next we show that $\beta^* = 0$. The relation between a^{j_1} and a^{j_2} means

$$\langle \pi^*, a^{j_2} \rangle + \beta^* = \langle \pi^*, \alpha a^{j_1} \rangle + \beta^* = \alpha \langle \pi^*, a^{j_1} \rangle + \beta^* = 0 = \langle \pi^*, a^{j_1} \rangle + \beta^*, \qquad (B.2)$$

implying

$$\alpha \langle \pi^*, a^{j_1} \rangle - \langle \pi^*, a^{j_1} \rangle = (\alpha - 1)\langle \pi^*, a^{j_1} \rangle = 0 \qquad\qquad (B.3)$$

and since $\alpha \neq 1$,

$$\langle \pi^*, a^{j_1} \rangle = 0. \qquad\qquad (B.4)$$

Hence, since $0 = \langle \pi^*, a^{j_1} \rangle + \beta^* = \beta^*$, $\beta^* = 0$. This establishes the existence of a strictly positive vector in $\Re^m$; namely, π^*, such that $\langle \pi^*, a^j \rangle \leq 0$; $a^j \in \mathcal{A}$, and $\langle \pi^*, a^0 \rangle = \langle \pi^*, a^{j_1} \rangle = \langle \pi^*, a^{j_2} \rangle = 0$. This makes the three points, a^{j_1}, a^{j_2}, a^0 efficient with respect to the constant returns DEA model.∎

So, to to check if a proportional pair should be considered in the construction of $\mathcal{F}^1$ simply score its midpoint ($\gamma = 1/2$ in the lemma). If the variable returns score yields an efficient point then the pair is eligible. Otherwise, neither point in the pair will belong to $\mathcal{F}^1$. Since this scoring can be done using the frame $\mathcal{F}^2$ this may be quite efficient if the frame is relatively small. If a pair of proportional points exist which are both efficient for the constant returns model, then we select either arbitrarily and assign it to $\mathcal{F}^1$.

We now present a modification of procedure **AllFrames** to handle the possibility of proportional points in the data domain.

$$\boxed{\begin{array}{c} \text{Procedure } \textbf{AllFrames.} \\ \textit{(Modified to include proportionality.)} \end{array}}$$

ASSUMPTION. There are no duplicate points in the data set $\mathcal{A}$.

PHASE 1. Identify the following four frames of the DEA data set $\mathcal{A}$.

<u>**Step 1.**</u> $\mathcal{F}^2 \leftarrow \text{VRFRAME}(\mathcal{A})$.

<u>**Step 2.**</u> $\mathcal{F}^3 \leftarrow \text{IRFRAME}(\mathcal{F}^2)$.

<u>**Step 3.**</u> $\mathcal{F}^4 \leftarrow \text{DRFRAME}(\mathcal{F}^2)$.

<u>**Step 4.**</u>

a. Find all K eligible proportional pairs, $\mathcal{M}^k$; $k = 1, \ldots, K$, of points in $\mathcal{F}^2$. That is, all points in $\mathcal{F}^2$ that are:

  $i)$ pairwise proportional, <u>and</u>

  $ii)$ both efficient with respect to the constant returns model.

b. Denote by a_+^k and a_-^k the largest and smallest points, respectively in pair $\mathcal{M}^k$ for $k = 1, \ldots, K$. Then, from $k = 1$ until $k = K$:

$$\mathcal{M}^+ \leftarrow a_+^k.$$

<u>**Step 5.**</u> $\mathcal{F}^1 \leftarrow (\mathcal{F}^3 \cap \mathcal{F}^4) \cup \mathcal{M}^+$.

PHASE 2. Select model and orientation and score DMUs using appropriate frame.

END PROCEDURE AllFrames.

Observations.

1. The no duplication assumption required for this modification of **AllFrames** is a relaxation of the non proportionality assumption originally required for the procedure. However, it is necessary to assure uniqueness of frames. A preprocessor may be applied to the data domain to check for, and remove, duplicates.

2. The main modification of procedure **AllFrames** needed to accommodate the possibility of proportionality is the addition of a new step after Step 3. The purpose of Step 4 is to identify any and all points in the variable returns frame, $\mathcal{F}^2$, that have an impact on the construction of the constant returns frame. These are only the proportional points which necessarily come in pairs since at most two proportional points can be extreme-efficient in $\mathcal{F}^2$. An additional restriction is that the proportional pair must both be efficient with respect to the constant returns model. This can be checked by scoring the midpoint of the pair using the variable returns model and invoking the lemma. The choice of orientation, if any, is irrelevant. If the midpoint is efficient the pair is eligible. Otherwise, they are excluded from the list $\mathcal{M}^k$.

3. The set $\mathcal{M}^+$ contains the the largest of the two points in each eligible proportional pair, $\mathcal{M}^k$, after the operations in Step 4b are completed. Thus, the cardinality of $\mathcal{M}^+$ is K. Selecting the largest from each pair is just one way to unambiguously allocate exactly one element from every, $\mathcal{M}^k$; $k = 1, \ldots, K$, for inclusion in $\mathcal{F}^1$ in Step 5. Actually, any rule can be used as long as precisely one element from each $\mathcal{M}^k$ ends up in $\mathcal{F}^1$.

4. The last step of the first phase of procedure **AllFrames** constructs the constant returns frame using $\mathcal{F}^3$ and $\mathcal{F}^4$. In the presence of proportionality, the intersection, $\mathcal{F}^3 \cap \mathcal{F}^4$, is just a subset of $\mathcal{F}^1$ and will exclude points which are part of proportional pairs when both are efficient with respect to the constant returns model. These excluded points constitute the set $\mathcal{M}^+$.

Conclusion. Perfect proportionality of data points is a rare event if the data comes from a real application or is generated randomly using sufficient precision. Moreover, proportionality has an impact on the outcome of procedure **AllFrames** only in the event when proportional pairs are eligible; that is, both efficient with respect to the constant returns model. This excludes proportional points that are inefficient or the midpoint of which is not in $\mathcal{E}^2$. For this reason, we

believe, the safeguards proposed here for this event will rarely have a significant impact on the final composition of $\mathcal{F}^1$.

Appendix C. Situations when using the frame provides sufficient conditions for identifying weak efficient DMUs. Solving DEA LPs composed only of the frame of the production possibility set to score DMUs reduces the ways the analysis can be confounded by inconclusive optimal solutions in oriented forms[†]. This situation is illustrated in the following example and accompanying figure. Consider the following five data points given in Table 2.

<div style="border: 1px solid black; padding: 10px;">

Table 2. DEA Data for Output-Oriented Constant Returns model.

</div>

DMU $\rightarrow$	1	2	3	4	5
Input 1	1	1	1	1	1
Output 1	7	10	13	1	4
Output 2	10	8	4	10	10

The analysis will use the output orientation and the "constant returns" assumption about the technology. It is clear after a simple inspection that DMUs 4 and 5 are inefficient in comparison with DMU 1. Suppose DMU 5 is scored using the standard approach; that is, by solving the following LP.
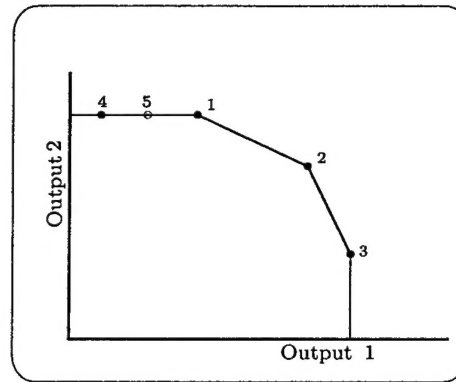
$$\max \quad \phi$$
$$\text{s.t.} \quad \lambda_1 + \lambda_2 + \lambda_3 + \lambda_4 + \lambda_5 \leq 1$$
$$7\lambda_1 + 10\lambda_2 + 13\lambda_3 + \lambda_4 + 4\lambda_5 - 4\phi \geq 0$$
$$10\lambda_1 + 8\lambda_2 + 4\lambda_3 + 10\lambda_4 + 10\lambda_5 - 10\phi \geq 0$$

$$\phi \text{ unrestricted, } \lambda_j \geq 0; \quad j = 1, 2, 3, 4, 5.$$

There are several optimal solutions for this LP. One is where $\lambda_1 = \lambda_4 = 1/2$ with all slacks zero. In this solution, the dual multiplier associated with the first output is zero.

Figure 3 depicts a configuration of data points that generates the situation in the example above. The plot represents a level set of the production possibility set parallel to the plane of the two outputs. The points $1, \ldots, 5$ correspond to the location of DMUs 1 through 5, respectively. We

---

[†] Unoriented forms where the objective function of the envelopment LP is the minimization of combinations of slack do not exhibit problems with "weak" efficiency.

Figure 3. Level surface on the output plane for data in Table 1.

can see from the figure that DMUs 4 and 5 are weakly efficient since they are on the boundary of the production possibility set but are clearly not as efficient as DMU 1. When scoring DMU 5, one solution combines DMUs 1 and 4 without slacks. However, since DMU 5 is on the "horizontal" part of the production possibility set, the multiplier for the first output must be zero. Based strictly on this LP solution, further tests would be required to resolve the status of DMU 5.

The frame of the production possibility set for this example is composed of the first three DMUs. If scoring proceeds using the frame, the score for DMU 5 results from the solution of the envelopment LP with only the first three variables plus the radial distance measure, ϕ. Clearly, the weakly complementary solution above is excluded. In fact, there is a unique optimal solution using the frame in this case; it is the one where $\lambda_1 = 1$ and the slack for the first output is 3. This solution satisfies the sufficient condition to conclude that DMU 5 is weakly efficient.

The general situation illustrated by the example is that of DMUs on a face of the production possibility set which is not on the efficient frontier. The situation arises if the face contains two or more weakly efficient DMUs such that one of them can be expressed using (in accordance with the oriented DEA model being used) other weakly efficient DMUs in lieu of slacks. Such a solution would be alternative to another solution involving only the extreme efficient DMUs combined with nonnegative values for some slacks. If the frame is used in the formulation of the LPs no weakly efficient DMU could be involved in scoring any other DMU and therefore these particular ambiguous solutions are excluded.

REPORT DOCUMENTATION PAGE

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Washington Headquarters Service, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188) Washington, DC 20503.
PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT DATE	3. DATES COVERED *(From - To)*
5/20/99	Interim	4/1/98 –5/31/99

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
A new framework for large scale Data Envelopment Analysis (DEA)	
	5b. GRANT NUMBER N00014-98-1-0766
	5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)	5d. PROJECT NUMBER
Jose H. Dula	98PR06739-00
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)	8. PERFORMING ORGANIZATION REPORT NUMBER
University of Mississippi	

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
Office of Naval Research Regional Office, Atlanta Suite 4R15, 100 Alabama Street NW Atlanta, GA 30303-3104	ONR
	11. SPONSORING/MONITORING AGENCY REPORT NUMBER

12. DISTRIBUTION AVAILABILITY STATEMENT

13. SUPPLEMENTARY NOTES

14. ABSTRACT
This project proposes to investigate a new framework for large scale DEA studies that will reduce computation times and increase analysis flexibility.
The methodology is based on an initial extraction of essential elements known as "extreme-efficient" DMUs. With this information, all four standard DEA models can be derived efficiently. The idea was tested on 10,500 enlisted men in the E-8 and E-9 pay grade category in a DEA model to assess how effective these men are in transforming their education and skills into performance measures valued by the Navy.

15. SUBJECT TERMS
Data Envelopment Analysis, Large scale performance evaluation

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			
					19b. TELEPONE NUMBER *(Include area code)*